Parallel Matching System for Digital Non text Information (Fingerprint Image) (Vot: 71739)

Project Leader:

Mazura Mat Din

Project Members :

Associate Professor Dr. Mohd Aizaini Maarof
Associate Professor Mazleena Salleh

2006

ABSTRACT

Ever since we were born, there are few parts of our body have been developed uniquely to represent our identities. Parts like retina patterns, veins, DNA, fingerprints are all unique for every human being. In order to recognize these unique patterns, scientists have spent billions of dollars and years of research and development to produce tools and techniques to identify humans based on their unique body parts. Today, the most highly utilized parts of our body for the identification purposes are fingerprints. Although there are so many readers and applications have been made for the job, there are still weaknesses that need to be addressed especially in the area of fingerprint matching. Therefore, the aim of this research is to find ways to increase the performance of fingerprint matching in response to the increased number of fingerprint records available in the system. This research focuses on developing a parallel matching process in a computer cluster to increase the performance of fingerprint matching, without taking into account additional features such as high throughput and others. The main idea is to minimize processing time with multiple processes executing parallels in several computers. As a proof of concept, a prototype consists of client and server processes has been developed. An interface which been integrated with a fingerprint device on client side enables a request (a fingerprint image) been send to a server process. Upon receiving a fingerprint image, the server process will place the request into a queue before distributing to a cluster of matching server throughout the network. The processing time for fingerprint matching has been improved especially when the amount of records increased. The detailed results of identification, duration of matching and similarity are discussed.

ABSTRAK

Manusia dilahirkan dengan memiliki anggota unik sebagai pengenalan dirinya. Antara anggota yang yang dikenalpasti dan digunakan ialah corak retina mata, pembuluh darah, DNA, dan cap jari di mana setiap satunya unik bagi seseorang individu. Saintis dari pelbagai bidang telah menjalankan pelbagai kajian dan teknik penyelidikan dalam membangunkan alatan untuk pengkelasan dan pengecaman identiti setiap individu manusia berdasarkan anggota badan berkenaan. Antara anggota berkenaan cap jari telah dikenalpasti sebagai satu mekanisme yang meluas digunakan. Walaupun terdapat pelbagai aplikasi dan alatan pengimbas cap jari, masih terdapat kelemahan di dalam proses pengecaman ini terutamanya jika melibatkan padanan cap jari dalam kuantiti yang banyak. Oleh itu, matlamat kajian ini adalah untuk mmeperbaiki masa tindak balas pemadanan imej cap jari dengan jumlah rekod cap jari yang bertambah di dalam sistem. Penyelidikan ini adalah tertumpu kepada pembangunan sistem selari yang melakasanakan pemadanan cap jari di dalam persekitaran komputer berkelompok untuk meningkatakan prestasi. Ini bertujuan untuk meminimumkan masa pemprosesan dengan mengimlimentasikan pemprosesan selari ke atas beberapa komputer. Prototaip yang terdiri dari proses pelanggan dan pelayan dibangunkan di samping antaramuka yang dintgrasikan bersama peranti imbasan cap jari. Imej cap jari yang diimbas akan dihantar ke adalam satu baris gilir sebelum diagihkan ke pelayan pemadanan melalui rangkaian. Dengan perlaksanaan prototaip ini, masa pemprosesan untuk pemadanan cap jari dapat diperbaiki terutamanya apabila pertambahan rekod berlaku. Hasil terperinci seperti hasil pengecaman, tempoh dan julat pemadanan dibincangkan lebih lanjut di dalam laporan ini.

ACKNOWLEDGEMENT

We would like to extend our appreciation to Research Management Center, Universiti Teknologi Malaysia for funding this project.

1.0 Introduction

Early studies in this research is to enhance the performance of processing the digital non-text information (in our case study is fingerprint image) by introducing a parallel compression system. During the process we found out that the results based on the early studies has already been achieved by using a compression tools available with less hassle. Therefore, we divert our application development towards the unsolved research on fingerprint matching processes but still maintaining the same method of solution by doing a parallel processing for the data. Details of this research will be further explained throughout every modules in this report.

High Performance Computing (HPC) generally defined as the technology that is used to solve computational problems that need significant processing power and need to quickly access and process very large amounts of data.

The ultimate goal of HPC is to reduce the time of execution for a computational or data-intensive application which allows businesses and researchers to solve difficult computational problems and do it faster than ever before. HPC is used extensively today in several areas such as financial modeling, computational fluid dynamics, and biomedical research also foresee suitable for digital non-text information such as fingerprint image that been selected as a data in this project.

At present, the most common HPC programming strategy is the use of parallelism. Parallelism can be achieved either by the use of multiple processors in a single computer or by the use of a set of processors across several computers (clusters). Based on HPC taxonomy (Microsoft Corp., 2001), A parallel high performance computing (HPC) introduced in this research were able to do fingerprint recognition on different fingerprint record set on several compute nodes.

2.0 Problem Statement

Nowadays, automatic fingerprint matching is becoming increasingly popular in systems, which control access to physical locations, computer or network resources, bank accounts, or register employee attendance time in enterprises. Current system in fingerprint matching only involved a stand alone application in a single computer. Combine with a capability of the matching algorithms this system only manage to works well for a small amount of data which is not applicable with the exponential growing of population today.

Give a pair of fingerprint images, the purpose of fingerprint identification system (FIS) is to decide either the fingerprints are identical, where the submitted fingerprints and candidate fingerprint are matched, or they are non-identical, where the two images are not matched. Current FIS components are developed by different people in a different period of time. Although each component of the FIS system works well, some difficulties still make doing test inconvenient and tedious. The challenges associated with the FIS is:

i. Performance: involve in fingerprint comparison to achieve a reasonable response time
ii. There is no graphics user interface for the FIS system
iii. Fingerprint image database takes too much space in the workstation doing FIS experiments

In this research, an automatic parallel fingerprint identification system in a cluster computer is developed to solve these problem and to assist the experiment of FIS.

3.0 Project Purpose

The project was aimed at producing an application that can enhance the performance of fingerprint identification and matching process. This is done through the used of parallel processing in a cluster computers.

4.0 Objective

The objectives of this project are to:

i. build an application and prototype that can help in identifying the fingerprint image in lesser time.
ii. and to study a parallel computer design that will enhance the performance of the matching process.

5.0 Scope of Study

To achieve the above objectives, below is the scope of this project:

i. To enable the non-text image registration or collection using a fingerprint terminal.
ii. To provide an identification server that host a server Development of the Identification Server
iii. To setup a cluster of identification nodes for parallel processing or matching.
iv. Development of the Fingerprint Image Database to host a sample of non-text information.
v. 8000 samples of fingerprint image were used in this prototype.

6.0 Parallel Cluster Computers Prototype

A cluster is a group of independent computer systems and thus forms a loosely coupled multiprocessor system. A network is used to provide inter-processor communications. Applications that are distributed across the processors of the cluster use either message passing or network shared memory for communication. A cluster computing system is a compromise between a massively parallel processing system and a distributed system .

The prototype consist of four nodes which separated by its functions. The building block of the architecture is the client node, a server or manager, and two fingerprint identification nodes that connected through local area network. The communication process between client process and server process, and server process and fingerprint identification process done via .NET Remoting infrastructure. The architecture is shown in Figure 1.

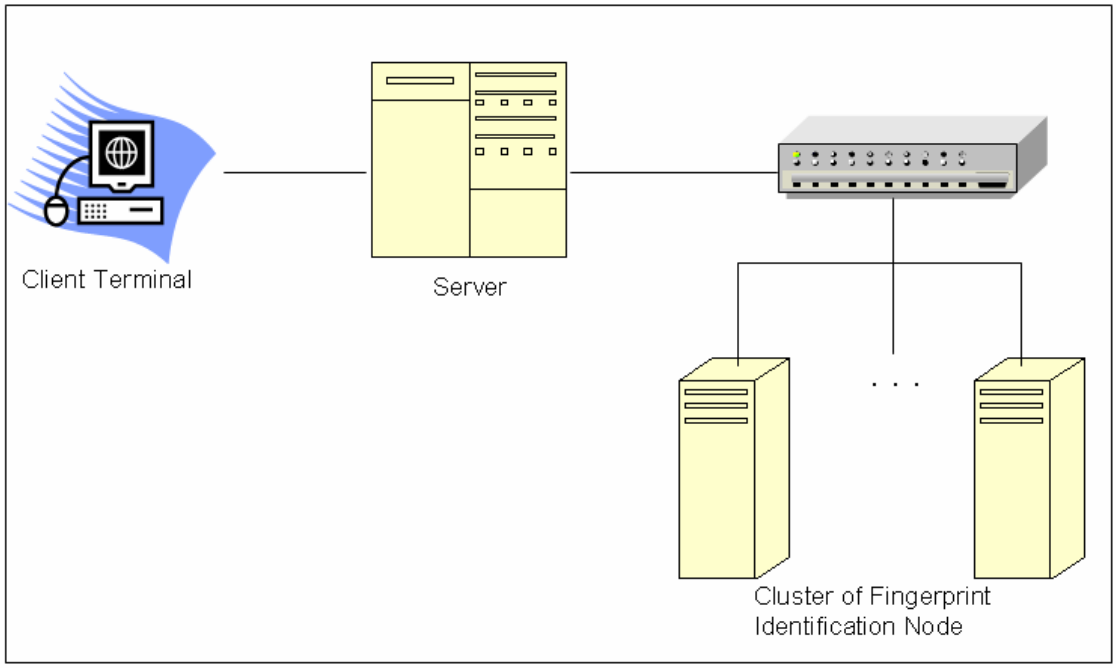

Figure 1: The System Architecture

i. Client node: it's the user interface; all the operations are performed from here. This is a place to submit user's testing samples and trigger the specific experiment on the server. The fingerprint image input is through fingerprint scanner or in a bitmap (BMP) file format.

ii. Server node: This is where the real work for the FIS happened. The server software will trigger a multicast datagram to specified IP and port number of fingerprint identification nodes. After receiving an image from client, it will invoke the FIS engine, pass the job/task to the FiD nodes and pass the results back

to clients. The distribution of fingerprint record will be done at server side. It is also responsible for accessing the fingerprint database.

iii. FiD Nodes: The server will update the node to become an active FiD node once it received a request from the corresponding nodes. Every request from the client will be inserted into a queue at server node before a balanced load of record is distributed among the active FiD nodes.

6.1 Development of the Fingerprint Terminal

The main task of the fingerprint terminal is to provide a user interface where end users will input fingerprint images. Fingerprint images could be scanned via attached fingerprint scanner or selected from existing bitmap files. Fingerprint identification requests will be sent to the identification server, and responds will be sent back to the terminal after the server finished processing the request sent from user.

Each terminal can communicate with the server. Communication is done by taking advantage of Microsoft .NET Remoting architecture to simplify development. Images are sent to the server as raw image format, consisting of an array bytes, with each byte holding the grayscale value of one pixel. Pixels are converted row-by-row, from top to bottom, left to right for each row.

While the server processes the request, the terminal will wait and hold off any other request while keeping track of the time elapsed. The terminal user interface will display the results and the elapsed time after receiving response from the server.

6.2 Development of the Identification Server

When user interacts with the fingerprint terminal, requests will be sent to the identification server to complete specific fingerprint identification jobs. Identification server works as a middleware for the fingerprint terminal to access identification nodes.

The identification server provides the following functions:

i. Evenly distribute fingerprints from database to identification nodes.
ii. Provides services for many terminals by enqueing their requests if there are multiple simultaneous requests.
iii. Distribute requests from terminals to identification nodes.
iv. Collect status from each identification nodes.
v. Halt identification process and prepare identification nodes for next available requests upon successful identification of currently executing identification process.
vi. Notify terminals when results are ready.

6.3 Development of the Identification Node

The actual identification process is executed at identification nodes. During loading of the application, each node will register itself to the identification server. Once prepared, the identification server will evenly distribute fingerprint records from the database to each identification node.

Each node will hold a unique set of fingerprint records to be matched to when receiving requests from the identification server. All records will be loaded into memory prior to the identification process to ensure highest performance possible, by reducing the overhead of disk access time.

6.4 Development of the Fingerprint Image Database

The fingerprint image database holds the data files and the corresponding properties for specific fingerprint. All the fingerprint records will be distributed by the identification server to all identification nodes.

There is only one table (Table 1) of interest, named ‘fingerprints’, which provides the combined data from several other tables. The attributes in this table are kept to a minimum, consisting of basic required information to identify each fingerprint records.

Table 1: Fingerprints Table

Attribute Names	Data Type	Size
id (primary key)	int	4
info	varchar	30
Features	varbinary	500

6.5 Performance Testing

Test was done against the system to collect performance data in various configurations – different number of records and identification nodes. The configurations used are as shown in Table 2.

Table 2: Test Configurations

Configuration	Number of Random Records	Number of Identification Nodes
1	30,000	1
2	30,000	2
3	30,000	3
4	100,000	1
5	100,000	2
6	100,000	3
7	200,000	1
8	200,000	2
9	200,000	3
10	300,000	1
11	300,000	2
12	300,000	3
13	400,000	1
14	400,000	2
15	400,000	4

There are four different test records that are inserted inbetween the random records. Each test record are labeled ‘Amry LT’, ‘Amry LI’, ‘Amry RT’ and ‘Amry RI’. For each test configuration, test records are inserted at the end of the first, second, third and forth quarter. Each test configuration are done twice where the mean time for identification processes are recorded.

6.6 Test Results

The test result was visualized in a graph form and been explained as follows:

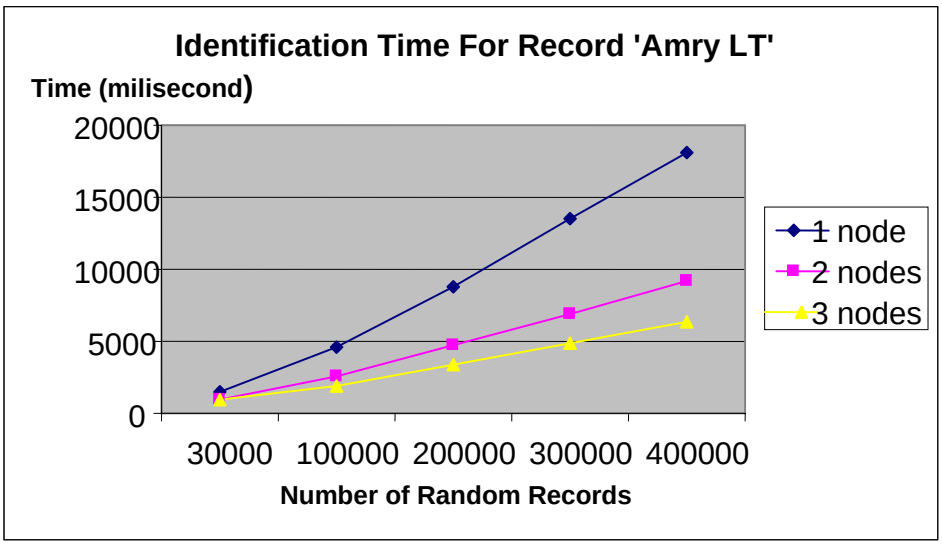

Figure 2: Identification Time for Record 'Amry LT'

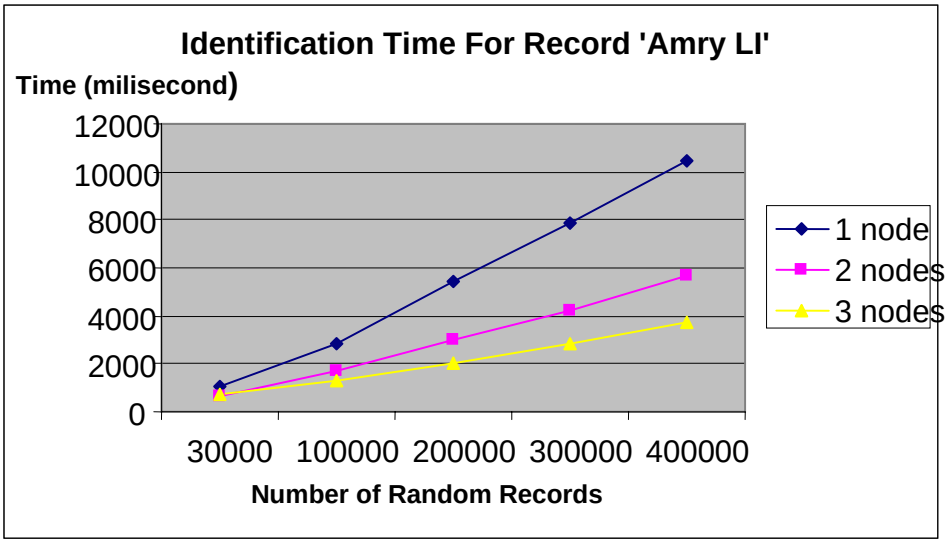

Figure 3: Identification Time for Record 'Amry LI'

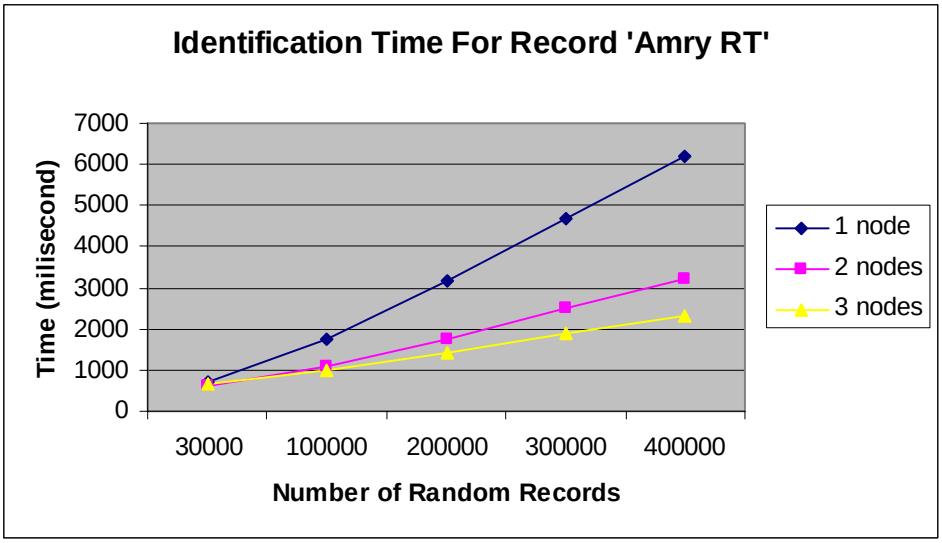

Figure 4: Identification Time for Record 'Amry RT'

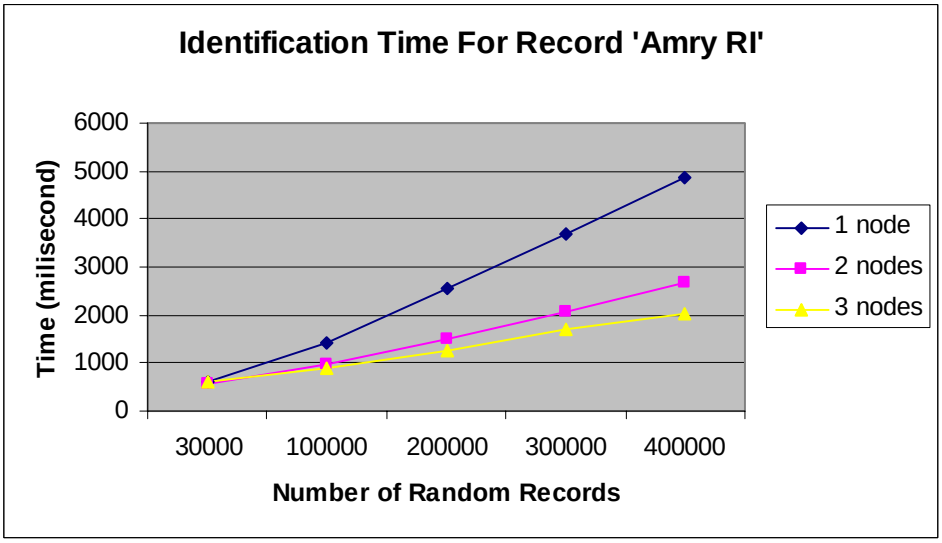

Figure 3: Identification Time for Record 'Amry RI'

From Figure 2, Figure 3, Figure 4 and Figure 5 above, the analysis graphs as follows:

i. The time it takes to return identification results will be reducing when there are more identification nodes used in the process.

ii. For tests which had less than 30,000 random records used, the usage of more than one identification node will not affect much. In most cases, it will even be slower since the communication overhead will take its toll.

iii. For tests which had more than 100,000 random records used, the graphs are progressing in a straight line.

7.0 Result

This project has produced a prototype application called APFIS (Automatic Parallel Fingerprint Identification Systems) that acts as a cluster of computers to enhance the matching processed for the fingerprint image. This prototype has been tested in Sekolah Kebangsaan Impian Emas, and proved that it can enhanced the process of matching the student fingerprint when using a cluster of computers compared to a single client-server matching. From the results produced, it is hoped that agencies such as Polis Di Raja Malaysia (PDRM), and forensics may also benefit from this application.

This application has been tested in Sekolah Kebangsaan Impian Emas and may be adopted by other schools including, and especially, the government's established smart schools through-out the country.

Paper that has been produced and presented in this project includes:

i. Mazura Mat Din, Ihsan Junaidi Ibrahim, "Parallel Compression System (PCS): A Review", COSTAM, Penang, 20-22 October 2001
ii. Mazura Mat Din, Adly Abd. Rahman, "Education Portal Berpandukan Pengesahan Menggunakan Teknik Fungsi Cincangan Dan Pangkalan Data Teragih" National Conference on Graphics and Multimedia 2002 (CoGramm), Malacca
iii. Poster presentation "Load Balancing Model Based on Loadserver Algorithm" di Research Seminar on RMK-7 & RMK-8, aerospace, Information Technology & Communication 2003.